Random Number Generators: Selection Criteria and Testing

Pierre L'Ecuyer and Peter Hellekalek

1 Introduction

A random number generator produces a periodic sequence of numbers on a computer. The starting point can be random, but after it is chosen, everything else is deterministic. The goal is to "fake" the realization of a sequence of independent and identically distributed (i.i.d.) random variables uniformly distributed over some set U, which is most often the unit interval $[0, 1]$. In the latter case, the random variables are denoted $U(0, 1)$.

In a more precise way, we define a *random number generator* (RNG) as a structure (S, U, μ, f, g), where S is a finite set of *states* called the state space, U is a set of *output* symbols called the output space, μ is a probability distribution over S, $f : S \to S$ is a mapping called the *transition function*, and $g : S \to U$ is another mapping called the *output function*. The *initial state* (or *seed*) s_0 is a random variable that follows the probability distribution μ. After s_0 has been chosen, the state evolves as $s_n = f(s_{n-1})$, $n \geq 1$, and the generator outputs $u_n = g(s_n)$ at step n. These u_n are the so-called *random numbers* produced by the generator. The sequence $\{s_n\}$ is eventually periodic with period length $\rho \leq |S|$, where $|S|$ denotes the cardinality of the set S. If the state is represented over b bits of memory, an upper bound is $\rho \leq 2^b$ and, for obvious reasons, we prefer ρ to be near 2^b. In this paper, we shall assume that the sequence is *purely periodic*, in the sense that the initial state s_0 is always revisited. In other words, the sequence has no transient part.

The goal is to make it hard to distinguish between the output of the generator and a typical realization of an i.i.d. uniform sequence over U. In *principle*, it is always possible to make the distinction, because the output sequence of the RNG is periodic. But if the period is very long (for example, more than 2^{100}) then no computer can exhaust it in our lifetime. So, there is still room for hope that one cannot distinguish the two sequences in *reasonable time*. For Monte Carlo simulation, we ask for less: we want the output sequence to behave like an i.i.d. uniform sequence *for practical purposes*, for the problem at hand.

It is common practice to select the parameters of RNGs according to certain *figures of merit*. For example, linear congruential or multiple recursive generators are usually selected via the so-called *spectral test*, which measures the "equilibrium" of their lattice structure. Tausworthe-type generators are selected by studying the equidistribution of their points in regular partitions of the unit hypercube into cubic cells of equal sizes. This gives rise to figures of merit measuring the equilibrium or equidistribution of all the points that can be produced by the generator over its entire period, and from all possible initial seeds in case it has more than one cycle. We refer to the contribution of the second author in this volume for further information on measures of equidistribution.

These figures of merit can be seen as *predictors* of the good or bad statistical behavior of an RNG. As a *general rule*, generators with long period and good figures of merit should pass the statistical tests, whereas those with bad figures of merit should fail "more easily". The aim of this paper is to examine this from the empirical viewpoint, for certain classes of generators commonly used in practice. For a number of selected empirical tests, we find (approximately) the minimal sample size required for the generator to clearly fail the test as a function of the size (i.e. period) of the generator and for different values of the figure of merit. We give an example of a test and an RNG family where the generators with the best figures of merit do *worse* in the tests than those with a not-so-good figure of merit. Nevertheless, the general rule given above still holds "typically".

In the next section, we discuss some principles underlying the construction of RNGs. The right approach is to design a generator based on theoretical considerations, keeping an eye of course on practical implementation issues, and test the generator empirically afterwards. Section 3 discusses empirical statistical testing for RNGs. Section 4 describes specific classes of tests: Disjoint and overlapping serial tests and tests based on close pairs of points in the unit torus. Section 5 defines certain collections of generators that will be used to make systematic experiments with some tests. Each such collection contains a generator of period length near 2^e for $e = 12, 13, \ldots, 34$. There are two collections of linear congruential generators (good and bad), one of explicit inversive generators, and one of combined cubic generators. Section 6 applies the tests of Section 4 to these collections of generators, in a systematic way. Section 7 summarizes what happens when they are applied to certain widely used generators.

2 Design Principles and Figures of Merit

2.1 A Roulette Wheel

An RNG can be viewed as a giant roulette wheel around which is written a sequence of output numbers. When the generator has more than a single cycle, we can view it as a collection of roulette wheels, one for each cycle. For simplification in the following discussion we assume that there is a single cycle, of length $\rho = |S|$. The arguments generalize easily to the case of several cycles. We also assume that $U = [0, 1]$.

Suppose that μ is the uniform distribution over S, and that the initial seed is random and uniformly distributed over the state space. If the generator is used to produce t output values, then this is equivalent to spinning the wheel to select a starting point at random and take the t successive values $u_0, \ldots, u_{t-1}$ that appear on the wheel from this starting point onward. What we want is for the vector $\boldsymbol{u}_{0,t} = (u_0, \ldots, u_{t-1})$ to be approximately uniformly distributed over the unit hypercube $[0, 1]^t$. This requirement captures both the uniformity and independence of the u_i.

Since s_0 is a random variable, $\boldsymbol{u}_{0,t}$ is a random vector, uniformly distributed over the multiset $\Psi_t = \{\boldsymbol{u}_{0,t} = (u_0, \ldots, u_{t-1}) : s_0 \in S\} = \{\boldsymbol{u}_{n,t}, 0 \leq n < \rho\}$, where the last equality holds only under our assumption that the generator has a single cycle of length $\rho = |S|$. Strictly speaking, Ψ_t is a *multiset* in the sense that repeated elements (if any) are counted as many times as they occur in the list. But we often just call it a *set*, for simplification. The aim is that the uniform distribution over Ψ_t be a good approximation of that over $[0, 1]^t$, for each t up to some reasonably large value. It is then natural to ask for Ψ_t to be very evenly spread over the unit hypercube, more uniformly than a typical sample of random points (i.e., *super-uniformly* distributed).

2.2 Sampling from Ψ_t

Suppose now that several t-dimensional points are needed. If one uses $\boldsymbol{u}_{n,t} = (u_n, \ldots, u_{n+t-1})$ for $n = 0, t, \ldots, (q-1)t$, say, then the concatenation of these points can be viewed as a single point in the qt-dimensional unit cube, and what we need in that case is good uniformity in the qt-dimensional space. But as qt gets higher, it becomes increasingly difficult in practice to analyze the uniformity of Ψ_{qt}. The points $\boldsymbol{u}_{0,t}, \ldots, \boldsymbol{u}_{(q-1)t,t}$ can be viewed as a "random" sample of q points taken uniformly from the set Ψ_t, without replacement. If ρ is several orders of magnitude larger than qt, there is practically no difference between *with* and *without* replacement. For further discussion about this argument, see [Com95, L'E97b, L'E98a].

A slightly different requirement would be to ask for the set Ψ_t to look like a typical set of i.i.d. $U(0, 1)$ random points. But what needs to behave like a set of i.i.d. uniforms is the set of points that are actually taken from Ψ_t by the random number generator, not necessarily the entire set. If the period is really large, as we recommend, the set Ψ_t will be so huge that only a negligible fraction of it can be generated in "feasible" time. Consequently, only a negligible fraction will be used. One could still argue that super-uniform generators tend to produce point sets that are so regular that some statistical tests (or some simulation problems) may be affected by the regularity for sample sizes that are relatively small compared with the period length. Indeed, we give an example later of a test that detects regularities in the "best" linear congruential generators for a sample size near the square root of the period length ρ. To escape the problem in this particular situation, ρ must be large enough so that generating $\sqrt{\rho}$ random numbers would take too much time to be feasible. If Ψ_t behaves more like a typical "random" point set, then a larger fraction of the period can be used and this is a (small) advantage.

In practice, the theoretical analysis of Ψ_t can be performed only for t up to a few dozens or at most a few hundreds, which is relatively small compared with the sample sizes that are commonly used in large simulations. And even if Ψ_t is well-distributed, this does not guarantee a good empirical behavior for the vectors $\boldsymbol{u}_{n,t}$, $n = 0, t, \ldots, (q-1)t$, when $q \ll \rho$. This is where empirical statistical testing comes into play: To assess the quality of the behavior of large segments of the generator's sequence, or of several successive points of the form $\boldsymbol{u}_{n,t}$, or of specific bits taken out of these numbers, or of certain particular transformations, and so on. Statistical tests are discussed in forthcoming sections.

To assess the super-uniformity of Ψ_t for large-period generators, generating the entire sequence of points is impossible, so one needs more sophisticated mathematical techniques. In practice, the techniques that are used are specific to certain classes of generators. The uniformity of Ψ_t is measured by *figures of merit* whose choice depends on practical considerations such as the ease of computing them.

In particular, the quality of linear congruential generators (LCGs) or multiple recursive generators (MRGs) is assessed by examining their lattice structure, most often via the spectral test, whereas linear feedback shift register generators are evaluated by measuring the equidistribution of their points in certain cubic boxes which form a partition of the unit hypercube. We give more details about these two special cases in what follows.

2.3 The Lattice Structure of MRGs

An MRG is defined by a linear recurrence of the form ([LBC93, L'E94, L'E96a, Nie92]):

$$x_n = (a_1 x_{n-1} + \cdots + a_k x_{n-k}) \bmod m; \tag{1.1}$$

$$u_n = x_n / m, \tag{1.2}$$

where m and k are positive integers called the modulus and the order, and $a_1, \ldots, a_k$ are integers between $-m$ and m, called the coefficients of the recurrence. The largest possible period for this recurrence is $\rho = m^k - 1$, attained if and only if m is prime and the characteristic polynomial $P(z) = z^k - a_1 z^{k-1} - \cdots - a_k$ is primitive over $\mathbb{Z}_m$, the finite field of residue classes modulo m. This can in fact be achieved with only two nonzero coefficients a_k and a_r, $1 \le r < k$. MRGs with very large (composite) moduli can be implemented indirectly by combining two or more MRGs with smaller moduli, as explained in [L'E96a, L'E97a]. When $k = 1$, the MRG becomes the classical LCG, in wide use since the fifties.

The set Ψ_t that corresponds to an MRG of the form (1.1–1.2) turns out to be the intersection of a lattice L_t with the half-open unit cube $[0,1)^t$ [Knu81, LBC93]. This implies that L_t is covered by each of several families of equidistant parallel hyperplanes. Among all such families that cover L_t (and also Ψ_t), choose one for which the distance between the successive hyperplanes is the largest, and let d_t be this distance. Computing this number d_t is often called the *spectral test* [Fis96, Knu81], for historical reasons. For a fixed cardinality of the set Ψ_t, a smaller d_t means that Ψ_t is more uniformly spread over the unit hypercube, and this is what we want.

Absolute lower bounds on d_t are available for general lattices having m^k points in the hypercube $[0,1)^t$. Let $d_t^*(m^k)$ be such a lower bound (for further details, see [Knu81] for $t \leq 8$ and [CS88, L'E98b, L'E97a] for more general t). The value of d_t can be normalized to

$$S_t = \frac{d_t^*(m^k)}{d_t},$$

which lies between 0 and 1. For any $\tau \geq k+1$, one can then define the worst-case *figure of merit*:

$$M_\tau = \min_{2 \leq t \leq \tau} S_t. \tag{1.3}$$

Searches for MRGs and LCGs with good figures of merit M_τ have been performed by several authors. Fishman and Moore [FM86] and Fishman [Fis90] found good LCGs with respect to M_6. L'Ecuyer [L'E98b] gives tables of good LCGs with respect to M_8, M_{16}, and M_{32}, for moduli ranging (approximately) from 2^8 to 2^{128}. A list of MRGs of different sizes, good with respect to a related criterion (the Beyer quotient) for all $t \leq 20$, is given is [LBC93]. L'Ecuyer [L'E97a] reports the results of extensive computer searches for good combined MRGs with respect to M_{32}. He also provides specific implementations for 32-bit and 64-bit computers. These generators are reasonably fast and their period lengths range from 2^{185} to 2^{1319} (approximately). We believe that they constitute excellent choices for general-purpose statistical or mathematical libraries, or for simulation software environments.

Other figures of merit can be defined based on the lattice structure analysis. For example, one can use quantities such as the shortest distance between any two points in the lattice, or the minimal number of hyperplanes that cover all the points of Ψ_t, or Beyer quotients, which are the ratio of the lengths of the shortest and longest vectors in a Minkowski-reduced basis of L_t, and so on. See, e.g., [AG85, Fis96]. The quantity d_t turns out to be the most popular.

2.4 Equidistribution for Regular Partitions in Cubic Boxes

Among the alternatives to (1.1–1.2), there is a large class of generators based on arithmetic modulo 2. Take again the recurrence (1.1) but define the output at step n by

$$u_n = \sum_{j=1}^{L} x_{ns+j-1} m^{-j}, \tag{1.4}$$

where s and L are two positive integers. This works for general m but for more concreteness let us assume henceforth that $m = 2$. This is the most popular case. This generator is called a *Linear Feedback Shift Register* (LFSR) or *Tausworthe* generator [Fis96, L'E94, Nie92, Tau65]. A commonly used and *economic* version is based on the recurrence $x_n = (x_{n-r} + x_{n-k}) \bmod 2$, whose characteristic polynomial is a trinomial. An efficient implementation can be found in [L'E96b, TL91] for the case where $s \le r$, $2r > k$, and k does not exceed the number of bits in the computer word length. It is widely accepted that LFSR generators should have a characterictic polynomial of large degree (e.g., $k > 100$ or so), whose coefficients behave somewhat like independent random bits (i.i.d. uniforms over the set $\{0, 1\}$) [Com91, Com95, MK96]. So, a rather simple figure of merit that first comes to mind for these generators is the number N_1 of coefficients a_i that are equal to 1. This number should be not too far from $k/2$. One way of obtaining efficient implementations of LFRSs with a lot of nonzero coefficients is by combining several ones by bitwise exclusive-or [TL91, L'E96b, L'E98c, WC93]. Another technique is the *twisted GFSR* (TGFSR) implementation [MK94, MN98].

Of course, the simple figure N_1 is far from sufficient to assess the quality of a generator. We also need some more direct measure of uniformity of the set Ψ_t. For LFSR generators, Ψ_t does not have a lattice structure in the real space as for MRGs. But LFSR generators have a lattice structure in a more abstract space, namely a space of formal series, and it turns out that this structure can be used to analyze the uniformity of Ψ_t in the unit hypercube via the notion of equidistribution. We now briefly explain this notion. For more details about its analysis and the relationship with the lattice structure in formal series, we refer the reader to [CLT93, L'E94, L'E96b, Tez95].

Partition the interval $[0, 1]$ into 2^ℓ segments of equal length, where ℓ is a positive integer such that $\ell \le L$. This determines a partition of the t-dimensional unit hypercube into $2^{t\ell}$ cubic boxes of volume $2^{-t\ell}$ each. Suppose that the multiset Ψ_t contains 2^k points. This multiset is called (t, ℓ)*-equidistributed in base* 2 if each box of the partition contains exactly $2^{k-t\ell}$ points. This happens if and only if the vector $(u_n, \ldots, u_{n+t-1})$, with each of its components truncated to the first ℓ bits, takes each of the $2^{t\ell}$ possible values the same number of times over the set Ψ_t. Of course, this can happen only for

$$\ell \leq \ell_t^* \stackrel{\text{def}}{=} \min(L, \lfloor k/t \rfloor).$$

In other words, the number of points cannot be less than the number of boxes. If Ψ_t is (t, ℓ_t^*)-equidistributed for all $t \leq k$, we call Ψ_t (and the generator) *maximally equidistributed* (ME). This is the best equidistribution that one can hope for. If Ψ_t satisfies the further requirement that when $\ell > \ell_t^*$ (i.e., more boxes than points) no box contains more than one point, then we call it (and the generator) *collision-free* (CF). ME-CF generators have their points sets (or multisets) Ψ_t very uniformly distributed into cubic partitions of the hypercube.

Based on this setup, one can define various figures of merit measuring the lack of equidistribution. For instance, let ℓ_t denote the largest ℓ for which Ψ_t is (t, ℓ)-equidistributed. This ℓ_t is called the *resolution* in dimension t, and $\Lambda_t \stackrel{\text{def}}{=} \ell_t^* - \ell_t$ is called the *resolution gap* in dimension t. From the viewpoint of equidistribution, generators with large resolution gaps are considered bad. The *maximal resolution gap*, $\max_{1 \leq t \leq k} \Lambda_t$, or the *total resolution gap*, $\sum_{t=1}^{k} \Lambda_t$, are examples of figures of merit that naturally come to mind in this context. For ME generators, both are equal to zero. Efficient methods for computing ℓ_t for LFSR generators are explained in [CLT93, L'E96b, Tez95].

The ME-CF property appears so strong that some may wonder if ME-CF generators exist at all. It turns out there are plenty of them among the LFSR generators and they are relatively easy to find. L'Ecuyer [L'E96b, L'E98c] gives several examples and concrete implementations, among certain classes of combined Tausworthe generators.

The TGFSR generator, in its original version, has very large resolution gaps. This was discovered by Tezuka [Tez95] and by its creators Matsumoto and Kurita [MK94], who proposed a modification that adds some *tempering* to the output. The idea is to multiply the bit vector of the output by some well-chosen matrix of bits, at each step, in order to improve the equidistribution. With the type of tempering that is used in [MK94, MN98], it has been proved that ME cannot be achieved. Nevertheless, these authors obtain large resolutions by choosing very large values of k (up to $k = 19937$ in [MN98]).

A notion stronger than that of (t, ℓ)-equidistribution can be defined as follows. Consider the family $\mathcal{P}_q$ of *all* partitions of the unit hypercube into 2^{k-q} so-called elementary dyadic boxes of volume 2^{q-k}, with the length of each side of each box equal to a non-negative power of $1/2$, and where all the boxes have the same dimensions and orientation. In other words, these boxes are of the form $\prod_{i=1}^{t} [a_i 2^{-g_i}, (a_i + 1) 2^{-g_i})$, for some integers a_i, $0 \leq a_i < 2^{g_i}$, $g_i \geq 0$, and $g_1 + \ldots + g_t = k - q$. The partition into cubic boxes discussed previously is one member of this family. The multiset Ψ_t is called a (q, k, t)*-net in base* 2 if each rectangular box of the partition contains exactly 2^q points, for each partition that belongs to $\mathcal{P}_q$. (Using a different notation than ours, this is usually called a (t, m, s)-net.) Niederreiter [Nie92] defines a figure of merit $r^{(t)}$ such that for all $t > \lfloor k/L \rfloor$, Ψ_t is a (q, k, t)-net in base 2 with $q = k - r^{(t)}$. Unfortunately, no efficient method is available for computing $r^{(t)}$ when t and k are large.

2.5 Other Measures of Divergence

The lattice structure analysis of MRGs via the spectral test and the test of equidistribution for LFSR generators are ways of assessing the discrepancy between the distribution of the multiset Ψ_t and the uniform distribution. There are of course several other ways of measuring this discrepancy.

Two forms of the rectangular discrepancy, often called simply *the discrepancy*, can be defined as follows. Suppose that Ψ_t contains N points. For any t-dimensional rectangular box of the form $R = \prod_{j=1}^{t}[\alpha_j,\ \beta_j)$ with $0 \le \alpha_j < \beta_j \le 1$, let $I(R)$ be the number of points $\boldsymbol{u}_n$ in Ψ_t that fall into R, and let $V(R) = \prod_{j=1}^{t}(\beta_j - \alpha_j)$ be the volume of R. Let $\mathcal{R}$ be the collection of all such boxes R, and define

$$D_N^{(t)} = \sup_{R \in \mathcal{R}} |V(R) - I(R)/N|.$$

This quantity is the so-called t-dimensional *extreme discrepancy* of Ψ_t. If we restrict $\mathcal{R}$ to those boxes which have one corner at the origin, the corresponding quantity is the *star discrepancy*, denoted $D_N^{*(t)}$. Other variants also exist, with different definitions of the family $\mathcal{R}$. See [Nie92] for further discussion. In all cases, *high* uniformity means *low* discrepancy.

To obtain a generator for which Ψ_t is super-uniform, one would look for the *smallest possible* discrepancy, Unfortunately, no efficient algorithm is available for computing $D_N^{(t)}$ or $D_N^{*(t)}$ for large N and moderate t. Lower and upper bounds are often available, but these bounds are usually expressed as asymptotic expressions as a function of N, and often as an average over a certain class of generators. On the other hand, the *discrepancy* is defined for all types of generators, not only for MRGs or LFSRs for instance.

Recently, a new family of equidistribution measures has been introduced by Hellekalek [Hel97], the *weighted spectral test*. Two realizations of this general concept are the classical diaphony (see [HN98]) and the dyadic diaphony [HL97]. These two figures of merit allow the same kind of number-theoretic estimates as discrepancy, but are much easier to compute. Furthermore, the asymptotic distribution of diaphony and dyadic diaphony has already been found, see Leeb [Lee97], Leeb and Hellekalek [LH98], and James, Hoogland, and Kleiss [JHK96]. Both quantities are related to mean-square integration errors over certain classes of functions and also to the so-called L^2-discrepancy, see Strauch [Str94]. We refer to the contribution of the second author in this volume for further information.

3 Empirical Statistical Tests

3.1 What are the Good Tests?

After designing a car on the computer, using all those nice mathematical tools, the car maker will try a prototype on the road. For RNGs, the *road test* means *some simulation problems*. A first thing to do is to try some simplified simulation problems, for which the answer is known. This is basically the idea of empirical statistical testing.

The statistical tests view the RNG as a black box. They make the hypothesis that the output is the realization of a sequence of i.i.d. $U(0,1)$ random variables. This is called the *null hypothesis*, denoted $\mathcal{H}_0$. One knows up-front that $\mathcal{H}_0$ is formally false. But the important question is "How can this affect our statistical results"?

Generally speaking, the answer is unknown. Heuristically, statistical tests try to catch up unusual departures from uniformity, or significant defects. The tests simply go fishing. Any finite set of tests can miss certain particular defects. No amount of testing can guarantee that a generator is foolproof.

In fact, any RNG has its own particular structure. As pointed out by Compagner [Com95], the bad things happen when the structure of the simulation problem at hand interacts constructively with the structure of the RNG. The difference between the good and bad RNGs is that the former have their structure *better hidden*, so that the problems that constructively interact with them are *harder to find* and occur less often in practice.

A *statistical test* can be defined by a real-valued function T of a finite (fixed or random) number of output values from the generator. The function T is called the *test statistic*. There is an infinite number of such functions. To apply the test, one needs to know (or have a good approximation of) the distribution function F of the random variable T under the null hypothesis $\mathcal{H}_0$. One can then compute the value t_1 taken by T and the corresponding p-value of the test, defined as $p_1 = P[T > t_1|\mathcal{H}_0] = 1 - F(t_1)$. For discrete distributions, the "$>$" is sometimes replaced by "$\geq$", depending on the context. If F is continuous, p_1 is a $U(0,1)$ random variable under $\mathcal{H}_0$.

When p_1 is extremely close to 0 or to 1, one concludes that the generator *fails* the test. In case of doubt, when the outcome is not clear enough, the test can be replicated independently until either failure becomes obvious or one can declare that the suspect p-value was obtained by chance. In some situations, the test may be only *single-sided* in the sense that failure is declared only if p_1 is too close to 0, or only if it is too close to 1. The choice of rejection area depends on what the test aims to detect. We will talk about a more elaborate concept, so-called *two-level tests*, later in this chapter.

Which tests are the good ones? The question has no general answer. It depends on what we want to use the RNG for. The relevant tests are those for which the distribution of the test statistic T under H_0 is very close to that of the random variable of interest in our simulation, or at least for which the structures that are likely to affect this random variable can be detected by the test.

But this is an elusive ideal, except perhaps for a few very important simulation problems for which people are ready to spend time and money to test the RNG they use for their particular problem. For the vast majority of simulation applications, the users cannot afford more than pick a "well-tested" RNG off the shelf. In most cases, they do not even have any control over the RNG, because it is hidden deep in the simulation software environment they use and they cannot change it. Moreover, even if one can test and choose the RNG, it is rarely easy (or even possible) to construct a test T known to be sensitive to the same type of structure as the simulation problem at hand. This is especially true in the case of very complicated stochastic models.

Therefore, we are back to heuristics. If both our simulation model and our RNG have (independently constructed) complicated structures, the chance that they interact constructively can be deemed negligible. The RNGs that are placed in general purpose software packages must be tested without knowing the millions of things they will be used for. These general purpose RNGs should be submitted to a wide variety of tests of different natures, able to detect different types of correlations or other "simple" structures likely to be harmful.

3.2 Two-Level Tests

Instead of just computing a single p-value p_1, one can replicate the test several times over disjoint parts of the sequence. If this is done N times, then one obtains N copies of T, say $T_1, \ldots, T_N$, the N corresponding p-values, say $p_{1,1}, \ldots, p_{1,N}$, and their empirical distribution $\hat{F}_N$. Under $\mathcal{H}_0$ and the additional assumption that T has a *continuous* distribution, $p_{1,1}, \ldots, p_{1,N}$ are i.i.d. $U(0,1)$. Note that this second assumption is important and often not satisfied in practice, because several commonly used test statistics have in fact a discrete distribution [Knu81]. But if this discrete distribution is well-approximated by a continuous one, the $p_{1,i}$ will be approximately $U(0,1)$. The hypothesis $\mathcal{H}_0$ can then be tested by measuring the *goodness-of-fit* between $\hat{F}_N$ and the uniform distribution. Standard goodness-of-fit tests such as those of Kolmogorov-Smirnov (KS), Anderson-Darling (AD), Cramer-von Mises, Watson, etc., can be used for that purpose. We suggest [Dur73, Ste86a] as good references about these types of tests.

3.3 Collections of Empirical Tests

A classical list of statistical tests for RNGs is given by Knuth [Knu81]. More recently, Marsaglia [Mar96] has made available a large battery of tests with specifically chosen parameters, under the name of DIEHARD.

This battery includes tests that are more "stringent" than the classical ones given in [Knu81], in the sense that more popular generators tend to fail them. Another large testing package called `TestU01` [L'E00] is in the work and will be made available sometime in the next century.

Certain types of tests simply throw n points in the t-dimensional unit hypercube and examine how they behave. These points are generally vectors $\boldsymbol{u}_i$ that belong to Ψ_t. A natural idea is to test the equidistribution of these points by partitioning the hypercube into boxes and counting how many points fall in each box. These (empirical) *equidistribution tests* have several variants and we examine some in Section 4.1. The name *serial test* often refers to them. Other types of tests are based on the distances between the close pairs of points in the hypercube. We look at some of them in Section 4.2.

There are of course many other types of tests, sensitive to different types of structures. For example, tests based on discrete *random walks* and certain *weight distribution tests* are generally good to detect correlations for certain families of generators such as GFSR or LFSR when their characteristic polynomials have few nonzero coefficients [MK94, SHB97, Tak94, VANK94, VANK95]. Various types of tests based on the notion of *entropy* are studied in [DR81, DMST95, LCC96, L'E97a, L'E97b, PK97]. Some of them are closely related to sum-functions of spacings, which define yet other classes of tests [L'E97c]. Among several additional references proposing specific tests or applying certain tests to random number generators, we mention [L'E92, LW97, Mar85, Ste86b].

4 Examples of Empirical Tests

4.1 Serial Tests of Equidistribution

Define a mapping $\iota : [0,1]^t \to \{0,\dots,k-1\}$ such that $\mathrm{Vol}(\iota^{-1}(j)) = 1/k$ for $j = 0,\dots,k-1$. This determines a partition of the unit hypercube into k pieces of equal volume. An easy way of doing the partition is by dividing the interval $[0,1]$ into d equal segments, which determines a partition of the unit hypercube into $k = d^t$ cubic boxes. Here, we shall consider only this type of partition, with the understanding that the techniques apply more generally.

Generate n disjoint vectors $\boldsymbol{u}_{ti,t} = (u_{ti},\dots,u_{ti+t-1})$, $i = 0,\dots,n-1$, and let X_j be the number of these vectors that fall into the jth box, for $j = 0,\dots,k-1$. Under $\mathcal{H}_0$, $(X_0,\dots,X_{k-1})$ is a random vector having the multinomial distribution with parameters $(n,\ 1/k,\ \dots,\ 1/k)$. (In this paper, the n serves both as an index for random numbers as in (1.1), and to denote the sample size in certain statistical tests as here. But the context indicates well the correct interpretation, so the reader should not be confused.)

As a measure of the distance (or *divergence*) between the empirical and uniform distributions, we consider test statistics of the form

$$Y = \sum_{j=0}^{k-1} f(X_j), \tag{1.5}$$

where f is a real-valued function. Different choices are available for f. For example:

$$\begin{aligned}
\text{Chi-square:} \quad & X^2 &=& \sum_{j=0}^{k-1} \frac{(X_j - n/k)^2}{n/k}, \\
\text{Negative entropy:} \quad & -H &=& \sum_{j=0}^{k-1} (X_j/n) \log_2(X_j/n), \\
\text{Collisions:} \quad & C &=& \sum_{j=0}^{k-1} (X_j - 1)\, I[X_j > 1], \\
\text{Empty boxes:} \quad & V &=& \sum_{j=0}^{k-1} I[X_j = 0] \;=\; k - n + C.
\end{aligned}$$

For all these examples, a larger value of the test statistic means that the points are *less* uniformly distributed in the boxes. The negative entropy $-H$ is in fact equivalent (via a linear transformation) to the *loglikelihood ratio* statistic G^2 for the multinomial distribution [LCC96]. Both G^2 and the Pearson chi-square X^2 are members of a parameterized class called the *power divergence* test statistics for the multinomial distribution, studied in Read and Cressie [RC88]. All other members of this class are also special cases of (1.5). The test based on X^2, with $n/k > 5$ or so, is often called the *serial test* or *t-tuple test* [Fis96, Knu81]. The Collision test is analyzed by Knuth [Knu81], who gives an algorithm to calculate the exact distribution of C.

Under the hypothesis $\mathcal{H}_0$, and under mild conditions on f (see [Hol72, LSW98, Mor75, RC88] for details), the following holds: If k is fixed and $n \to \infty$,

$$Y_C \stackrel{\text{def}}{=} \frac{Y - E[Y] + (k-1)\sigma_C}{\sigma_C} \Rightarrow \chi^2(k-1), \tag{1.6}$$

where $\sigma_C^2 = \mathrm{Var}[Y]/(2(k-1))$, $\Rightarrow$ means convergence in distribution, and $\chi^2(k-1)$ is the chi-square distribution with $k-1$ degrees of freedom. We call this the *dense case*. If $k \to \infty$, $n \to \infty$, and $n/k \to \delta$ where $0 < \delta < \infty$, then

$$Y_N \stackrel{\text{def}}{=} \frac{Y - E[Y]}{\sigma_N} \Rightarrow N(0,1), \tag{1.7}$$

where $\sigma_N^2 = \mathrm{Var}[Y]$ and $N(0,1)$ is the standard normal distribution. We call this the *sparse case*. The statistics X^2, $-H$, G^2, C, and V, mentioned above all satisfy the conditions in the sparse case. However, C and V *do not* satisfy the conditions for the dense case: In that case, V converges to 0 with probability one as $n \to \infty$. General expressions for the exact mean $E[Y]$ and the exact variance $\mathrm{Var}[Y]$ are given in [LSW98]. The expression for the variance involves $O(n^2)$ terms, but in the sparse case, only a few of these terms are non-negligible, so one can easily obtain a very close numerical approximation. In other cases, asymptotic values are available.

Since C and V are integer-valued random variables, the normal approximation is not so good for them when their expectation is too small (say, less than 50 or so). In that case, they are approximately Poisson. When $\lambda' = n/k \ll 1$, the probability that any given box is empty is approximately $e^{-\lambda'}$, so $E[V] \approx ke^{-\lambda'}$ and $E[C] \approx n + k(e^{-\lambda'} - 1) \approx n^2/(2k)$.

The vectors $\boldsymbol{u}_{ti,t}$ defined above are independent under $\mathcal{H}_0$. The *overlapping* vectors $\boldsymbol{u}_{i,t}$, $i = 0, 1, 2, \ldots$ are not independent. If we use them to define the X_j, the limit theorems for Y_C and Y_N above no longer apply and the analysis is more difficult than for the disjoint case, but still something can be done for certain properly-defined statistics. Let $X^{\o}_{t,j}$ be the number of overlapping vectors that fall into cell j and define

$$X^2_{(t)} = \sum_{j=0}^{k-1} \frac{(X^{\o}_{t,j} - n/d^t)^2}{n/d^t}. \tag{1.8}$$

Then

$$\tilde{X}^2 = X^2_{(t)} - X^2_{(t-1)} \Rightarrow \chi^2(d^{t-1}(d-1))$$

as $n \to \infty$ for d and t fixed. This was proved by Good [Goo53]. The standard rule of thumb recommends to take $n/k \geq 5$ for the chi-square approximation to be good. This is called the *overlapping serial test* or the *overlapping t-tuple test* (often with m instead of t in the notation) in the literature [Alt88, LW97, Mar85, Weg95].

In the sparse case, we believe that $\tilde{X}^2$ converge to a normal with mean $d^{t-1}(d-1)$ and variance $2d^{t-1}(d-1)$. No rigourous proof of this is currently available according to our knowledge but plenty of numerical experimentation supports it. This is also in accordance with the chi-square asymptotic distribution of $\tilde{X}^2$ in the dense case.

When $\lambda' = n/k$ is very small, the overlapping version of C is approximately Poisson with the same mean as in the non-overlapping case, namely $n^2/(2k)$. The overlapping version of V is heuristically discussed in [Mar85, MZ93]. In [MZ93], the authors claim that it is approximately normal with mean $ke^{-\lambda'}$ and variance $ke^{-\lambda'}(1 - 3e^{-\lambda'})$. This approximation is good when λ' is between 2 and 5 (approximately) but it quickly deteriorates when λ' moves away from this range [LSW98].

The interest of the overlapping tests comes from the fact that for a given sample size n, they require approximately t times less random numbers than their non-overlapping counterparts while being almost as effective when testing most commonly used RNGs. We will give examples of this later on in this paper.

4.2 Tests Based on Close Points in Space

We consider again the same set of non-overlapping vectors $\boldsymbol{u}_{ti,t} = (u_{ti}, \ldots, u_{ti+t-1})$, $i = 0, \ldots, n-1$, as in the previous subsection. Let $\|\cdot\|_p$ denote the L_p-norm in $[0,1)^t$ viewed as a *unit torus*, for any $p \in [1,\infty]$, and let V_t be the volume of the unit ball $\{x \in \mathsf{R}^t \mid \|x\|_p \le 1\}$. (We choose the torus to get rid of boundary effects.) For any $s \ge 0$, let $Y_n(s)$ be the number of pairs (i,j), with $i < j$, such that

$$\|\boldsymbol{u}_{tj,t} - \boldsymbol{u}_{ti,t}\|_p \le \left(\frac{2V_t s}{n(n-1)}\right)^{1/t}.$$

Under $\mathcal{H}_0$, for each $s_1 > 0$, the stochastic process $\{Y_n(s), 0 \le s \le s_1\}$ converges weakly to a Poisson process of unit rate when $n \to \infty$. See [LCS97, RS78, SB78] for details.

The spacings between the first few successive jumps of Y_n should be approximately i.i.d. exponentials with mean one, and this can be the basis of a test. More specifically, if $T_{n,0} = 0$ and if $T_{n,1}, T_{n,2}, \ldots$, are the jump points of Y_n, the spacings are $\Delta_{n,i} = T_{n,i} - T_{n,i-1}$, $i \ge 1$, and the random variables $W^*_{n,i} = 1 - \exp[-\Delta_{n,i}]$, $i = 1, \ldots, M$, are approximately i.i.d. $U(0,1)$ under $\mathcal{H}_0$ if n is large and M is relatively small. As a simple rule of thumb, it is reported in [LCS97] that the approximation is accurate for $n \ge 4M^2$. Values of M in the range from 10 to 100 are reasonable. Computing the M nearest pairs turns out to be faster for larger p and the supremum norm ($p = \infty$) is a very good choice. Values of n up to 10 millions or so can be handled if t is not too large (this depends mainly on the size of the available computer memory).

To test the goodness-of-fit of the empirical distribution of $W^*_{n,1}, \ldots, W^*_{n,M}$ to the uniform distribution, the Anderson-Darling (AD) test statistic happens to be an excellent choice. The reason is that for the generators whose structure is too regular, the jumps of Y_n tend to cluster and this produces several $W^*_{n,i}$ that are very close to zero. When this happens, their empirical distribution goes up quickly near the origin. The AD test statistic is quite appropriate to detect this type of behavior, because it is most sensitive to departures from the uniform distribution near the extremities of the interval $[0,1]$. The AD statistic here is:

$$\begin{aligned} A_M^2 &= \int_0^1 \frac{(\hat{F}(u) - u)^2}{u(1-u)} du \\ &= -M - \frac{1}{M} \sum_{j=1}^{M} \left\{(2j-1)\ln(U_{(j)}) + (2M+1-2j)\ln(1-U_{(j)})\right\} \end{aligned}$$

where $U_{(1)} \leq \cdots \leq U_{(m)}$ are $W^*_{n,1}, \ldots, W^*_{n,m}$ sorted by increasing order and $\hat{F}$ is their empirical distribution. The hypothesis $\mathcal{H}_0$ is rejected when A^2_M takes a value a that is too large; i.e., when its p-value $p_1 = P[A^2_M > a|\mathcal{H}_0]$ is too small. This is the *M-nearest-pairs test* (M-NP test) proposed by L'Ecuyer, Cordeau, and Simard [LCS97].

A simpler version of this test is proposed and studied by Ripley [Rip77, Rip87], using only $M = 1$. The null hypothesis is rejected when $W^*_{n,1}$ is too close to either 0 or 1. The p-value is simply $p_1 = W^*_{n,1}$. This is the *nearest-pair test* (NP test).

These two tests can of course be applied in a two-level setup as explained previously. Replicate them N times and compare the empirical distribution of the N p-values to the uniform distribution using a second-level AD test. For the two-level NP test, it is typically much better to apply a *spacings transformation* or a *power ratio transformation* to the N values of p_1 before applying the AD test [LCS97]. The reason is that the values of p_1 often tend to cluster, and these two transformations change the small spacings into values that are very close to zero or very close to 1. The empirical distribution of the transformed values then becomes a good prey for the AD test at the second level. See [LCS97] for more details. We give examples later on.

5 Collections of Small RNGs

5.1 *Small Linear Congruential Generators*

Table 1.1 gives a list of LCGs with prime moduli m of different sizes and period $\rho = m - 1$. These LCGs are defined by

$$x_n = a x_{n-1} \bmod m, \tag{1.9}$$
$$u_n = x_n / m, \tag{1.10}$$

where a is a primitive root modulo m. For each integer e from 12 to 36, we took the largest prime m less than 2^e and found one *good* and one *bad* LCG with that modulus. Here, *good* and *bad* should be understood to be with respect to the spectral test. The good LCGs are those with the best figure of merit M_8 that were found in [L'E98b], for each prime m. The bad ones have been chosen to have $S_2 = M_2 \approx 0.05$, so they have a mediocre lattice structure in 2 dimensions.

5.2 *Explicit Inversive Congruential Generators*

We have a collection of *explicit inversive congruential generators* (EICG) defined by

$$x_n = (an)^{-1} \bmod m, \tag{1.11}$$

TABLE 1.1. Some good and bad (baby) LCGs

m	a (good LCG)	a (bad LCG)
$2^{12} - 3 = 4093$	209	5
$2^{13} - 1 = 8191$	884	2341
$2^{14} - 3 = 16381$	572	2731
$2^{15} - 19 = 32749$	219	10
$2^{16} - 15 = 65521$	17364	17
$2^{17} - 1 = 131071$	43165	68985
$2^{18} - 5 = 262139$	92717	203883
$2^{19} - 1 = 524287$	283741	458756
$2^{20} - 3 = 1048573$	380985	213598
$2^{21} - 9 = 2097143$	360889	202947
$2^{22} - 3 = 4194301$	914334	4079911
$2^{23} - 15 = 8388593$	653276	2696339
$2^{24} - 3 = 16777213$	6423135	486293
$2^{25} - 39 = 33554393$	25907312	5431467
$2^{26} - 5 = 67108859$	26590841	42038579
$2^{27} - 39 = 134217689$	45576512	24990322
$2^{28} - 57 = 268435399$	31792125	31842465
$2^{29} - 3 = 536870909$	16538103	8903330
$2^{30} - 35 = 1073741789$	5122456	3930720
$2^{31} - 1 = 2147483647$	1389796	868723
$2^{32} - 5 = 4294967291$	1588635695	1123161
$2^{33} - 9 = 8589934583$	7425194315	1026767
$2^{34} - 41 = 17179869143$	5295517759	10045
$2^{35} - 31 = 34359738337$	3124199165	10052
$2^{36} - 5 = 68719476731$	49865143810	102254510

$$u_n \quad = \quad x_n/m, \tag{1.12}$$

where the inverse of an modulo m can be computed as $x_n = (an)^{m-2} \bmod m$. All the generators in our EICG collection have the same prime modulus m as the LCGs, their multiplier is $a = 123$, and their period length is m. This type of nonlinear generator has been proposed by Eichenauer-Herrmann and is discussed in [Eic92, EHW97, Hel95, LW97].

5.3 Compound Cubic Congruential Generators

Our next collection contains a recent type of generator, the *compound cubic congruential generators* (CCCG), in the form proposed by Eichenauer-Herrmann [EH97], with two components. These generators are defined as follows. Let m_1 and m_2 be two distinct primes, a_1 and a_2 two positive integers less than m_1 and m_2, respectively, and put:

$$x_{1,n} = (a_1 x_{1,n-1}^3 + 1) \bmod m_1, \quad (1.13)$$

$$x_{2,n} = (a_2 x_{2,n-1}^3 + 1) \bmod m_2, \quad (1.14)$$

$$u_n = (x_{1,n}/m_1 + x_{2,n}/m_2) \bmod 1. \quad (1.15)$$

For certain values of m_1 and a_1, the recurrence (1.13) has the maximal period length m_1, and similarly for (1.14). These values are not so easy to find, because simple maximum period conditions are not available, and because these values are rare. Fourteen pairs (m_1, a_1) that produce maximum period are given in [EH97], for m_1 near 2^{15}. Table 1.2 gives the pairs that we use to construct our collection of CCCGs. These pairs (m_j, a_j) all yield the maximal period m_j, and they were found by brute force computations. For each e, the period length of the compound generator is equal to $m = m_1 m_2$, which is close to 2^e.

These nonlinear generators tend to have a much less regular structure than the LCGs or MRGs when a significant fraction of the period length is used, just like the inversive ones, but they run much faster than the latter, almost as fast as the LCGs. They seem to offer a good compromise between (slow) inversive generators and (too regular) LCGs.

6 Systematic Testing for Small RNGs

This section reports on empirical experiments made with the tests and the RNG collections described in the two previous sections. For each collection and each type of test, we find (empirically) some function φ and some constant ν^* such that $n \approx 2^{\nu^*}\varphi(e)$ is a crude approximation of the minimal sample size n at which the tests decisively reject the generators. As a general rule, for the two types of LCGs, we take $\varphi(e) = 2^{e/2} \approx \sqrt{\rho}$, whereas for the inversive and the cubic generators, we take $\varphi(e) = 2^e \approx \rho$. These functions turn out to be the right choices for ν^* to be approximately independent of e. The tests were carried out with a preliminary version of the `TestU01` package of the second author.

6.1 *Serial Tests of Equidistribution for LCGs*

Our first set of tests is in 2 dimensions ($t = 2$). For all the generators, we run the tests with sample sizes (the number of points in the hypercube) $n = 2^{\nu}\varphi(e)$, for $\nu = -5, \dots, 8$. The number of boxes is $k = (\lfloor (k')^{1/t} + 1/2 \rfloor)^t$, where each of the following four rules is tried for the choice of k': $k' \approx 2^e$, $k' \approx 2^{e+3}$, $k' \approx 2^{e-3}$, and $k' \approx n/8$. The transformation from k' to k is simply to make sure that $d = k^{1/t}$ is an integer. We compute X^2, $-H$, and their p-values with the normal approximation for $k \geq n$, and with the chi-square approximation otherwise. We also compute C and its p-value, using

TABLE 1.2. Full-period compound cubic congruential generators

e	m_1	a_1	m_2	a_2
12	59	15	47	7
13	101	61	59	15
14	101	61	83	9
15	251	135	101	61
16	251	135	233	61
17	503	445	251	135
18	503	445	491	277
19	1019	437	503	445
20	1019	437	1013	31
21	2039	243	1019	437
22	2039	243	2027	349
23	4091	494	2039	243
24	4091	494	4079	3757
25	8147	2410	4091	494
26	8147	2410	8111	2257
27	16361	5595	8147	2410
28	16361	5595	16319	3013
29	32693	1190	16361	5595
30	32693	1190	32687	4535
31	65519	512	32693	1190
32	65519	512	65447	27076
33	131063	110230	65519	512
34	131063	110230	130859	48249

the Poisson approximation when $E[C] < 50$ and the normal approximation otherwise. In fact, when $E[C]$ is too small, the normal approximation gets bad not only for C, but for X^2 and $-H$ as well, as we shall see in a moment. More details about which distribution is most appropriate in which situation can be found in [LSW98].

The results are summarized in a series of tables that follow. In these tables, only the p-values less than 0.01 or larger than 0.99 are given. We call them the *suspect* p-values. The entries where the p-values are not suspect are left blank. The p-values smaller than 10^{-15} are denoted by ϵ. A negative entry $-x$ means a p-value of $1 - x$. In particular, $-\epsilon$ means $> 1 - 10^{-15}$. In general, the columns with a larger ν than those given in a table contain only $\pm\epsilon$ and the columns with a smaller ν than what appears in the table are mostly blank.

When the Poisson approximation is used for C, the p-value printed in the table is defined as follows:

$$p = \begin{cases} P[Y \geq c] & \text{if } P[Y \geq c] < P[Y \leq c]; \\ 1 - P[Y \leq c] & \text{otherwise,} \end{cases}$$

where c is the value of C and Y is a Poisson random variable with mean $\mu = E[C|H_0]$. With this definition, a value of p close to zero [resp., close to one] means that C took a much larger [resp., much smaller] value than expected.

Table 1.3 gives the p-values for X^2, for $k' = 2^e$. The value of ν where the suspect p-values start to appear, as well as the value of ν where the $\pm\epsilon$'s start to appear, are pretty much the same for all e. Let us define $\tilde{\nu}$ and ν^* as the smallest ν for which the majority of the p-values are suspect and the smallest ν for which the majority of the p-values are $\pm\epsilon$, respectively. In Table 1.3, $\tilde{\nu} = 2$ and $\nu^* = 4$. This means that all the good LCGs with period length ρ are definitely rejected by the sparse serial test based on X^2 with $k \approx 2^e$ and $n = 16\sqrt{\rho}$ points. With $n \approx 4\sqrt{\rho}$, the test starts to give a majority of suspect p-values.

TABLE 1.3. The p-values for Pearson's X^2, for $t = 2$ and $k \approx \rho$.

e	$\nu = 1$	$\nu = 2$	$\nu = 3$	$\nu = 4$	$\nu = 5$
12			–4.0E-004	–5.9E-014	$-\epsilon$
13				–3.2E-010	$-\epsilon$
14			–1.0E-004	–9.5E-012	$-\epsilon$
15		–6.7E-003	–5.0E-006	–7.5E-015	$-\epsilon$
16		–2.3E-003	–9.1E-007	$-\epsilon$	$-\epsilon$
17		–6.7E-003	–1.0E-004	$-\epsilon$	$-\epsilon$
18			–1.3E-003	–1.1E-013	$-\epsilon$
19		–6.7E-003	–1.5E-007	$-\epsilon$	$-\epsilon$
20		–2.3E-003	–3.7E-007	$-\epsilon$	$-\epsilon$
21		–6.7E-003	–2.0E-004	–7.6E-013	$-\epsilon$
22			–7.3E-004	–5.4E-010	$-\epsilon$
23		–2.3E-003	–9.1E-007	$-\epsilon$	$-\epsilon$
24		–2.3E-003	–2.1E-008	$-\epsilon$	$-\epsilon$
25			–9.1E-007	$-\epsilon$	$-\epsilon$
26		–6.7E-003	–4.9E-006	$-\epsilon$	$-\epsilon$
27		–6.7E-003	–4.9E-006	$-\epsilon$	$-\epsilon$
28			–4.0E-003	–7.3E-015	$-\epsilon$
29		–2.3E-003	–7.7E-009	$-\epsilon$	$-\epsilon$
30			–5.0E-005	$-\epsilon$	$-\epsilon$
31		–6.7E-003	–1.5E-007	$-\epsilon$	$-\epsilon$
32			–1.0E-004	–1.5E-014	$-\epsilon$
33			–2.4E-005	–3.1E-011	$-\epsilon$
34			–2.4E-005	$-\epsilon$	$-\epsilon$

Table 1.4 gives the p-values of the test statistic C, for the same setup. The expected number of collisions in these tests is approximately $2^{2\nu-1}$. Therefore, the Poisson approximation was used for $\nu \leq 3$ and the normal approximation for $\nu \geq 4$. The entries where the normal approximation was

TABLE 1.4. Some p-values for C (collision test), for $t = 2$ and $k \approx \rho$.

e	$\nu = 1$	$\nu = 2$	$\nu = 3$	$\nu = 4$	$\nu = 5$
12			–2.8E-004	$-\epsilon$	$-\epsilon$
13				–6.8E-011	$-\epsilon$
14			–2.7E-005	–1.9E-012	$-\epsilon$
15		–3.2E-003	–1.6E-007	–1.1E-015	$-\epsilon$
16		–3.5E-004	–5.6E-009	$-\epsilon$	$-\epsilon$
17		–3.1E-003	–2.0E-005	$-\epsilon$	$-\epsilon$
18			–7.3E-004	–6.4E-014	$-\epsilon$
19		–3.1E-003	–8.4E-011	$-\epsilon$	$-\epsilon$
20		–3.4E-004	–6.8E-010	$-\epsilon$	$-\epsilon$
21		–3.0E-003	–4.9E-005	–6.4E-013	$-\epsilon$
22			–3.0E-004	–5.0E-010	$-\epsilon$
23		–3.4E-004	–4.3E-009	$-\epsilon$	$-\epsilon$
24		–3.4E-004	–4.3E-013	$-\epsilon$	$-\epsilon$
25			–4.2E-009	$-\epsilon$	$-\epsilon$
26		–3.0E-003	–1.1E-007	$-\epsilon$	$-\epsilon$
27		–3.0E-003	–1.1E-007	$-\epsilon$	$-\epsilon$
28			–2.8E-003	–7.2E-015	$-\epsilon$
29		–3.4E-004	–1.3E-014	$-\epsilon$	$-\epsilon$
30			–5.6E-006	$-\epsilon$	$-\epsilon$
31		–3.0E-003	–7.6E-011	$-\epsilon$	$-\epsilon$
32			–1.7E-005	–1.5E-014	$-\epsilon$
33			–1.7E-006	–3.1E-011	$-\epsilon$
34			–1.7E-006	$-\epsilon$	$-\epsilon$

used are remarkably similar to the corresponding entries in Table 1.3. In fact, if we had used the normal approximation everywhere, all the corresponding entries in the two tables would be nearly identical. The difference between the two tables in the columns for $\nu = 2$ and 3 is almost entirely due to fact that we use a Poisson (instead of normal) approximation for the distribution of C for these values of ν. The Poisson approximation is quite accurate in this area, so we must conclude that the normal approximation is not very good.

For $-H$ the p-values are almost the same as those of X^2 and C with the normal approximation. Why are these values so similar? Since $n \ll k$ and $E[C]$ is quite small here, most boxes contain 0 or 1 point, a few boxes contain 2 points, and in most cases no box contains more than 2 points. When the latter happens, C is the number of boxes that contain 2 points and there is a one-to-one correspondence between C, X^2, and $-H$, so the p-values of these three statistics should be almost the same. Under $\mathcal{H}_0$, for $n \ll k$, the number of boxes where X_i exceeds 2 is approximately Poisson with mean $\lambda_2 = n(n-1)(n-2)/(6k^2)$ [AGG89].

Considering that the p-values of these three statistics are almost the same when their distribution is approximated by the normal, and that they should effectively be nearly the same under the exact distributions, it follows that when the normal approximation is bad for C because $E[C]$ is too small, it is also bad for X^2 and $-H$. For instance, the correct p-values of X^2 in Table 1.3 should be much closer to the values of Table 1.4. Most of the difference in the two middle columns ($\nu = 2$ and 3) is caused by the lack of normality of X^2 for $\nu \leq 3$.

The $-\epsilon$ entries in the tables mean that X^2, $-H$, and C are much smaller than expected; i.e., that almost all of the points fall into different boxes. This is a consequence of the regular structure of the LCGs.

TABLE 1.5. The p-values of C for the bad LCGs, for $t = 2$ and $k \approx \rho$.

e	$\nu = -3$	$\nu = -2$	$\nu = -1$	$\nu = 0$	$\nu = 1$
14		4.5E-004	7.0E-003	3.1E-009	ϵ
15			9.2E-006	1.6E-010	ϵ
16		4.6E-004	7.1E-003	6.0E-008	ϵ
17		4.8E-004		6.0E-008	ϵ
18			7.1E-003	3.4E-009	ϵ
19		4.7E-004		3.4E-009	ϵ
20				1.7E-004	ϵ
21				3.4E-009	ϵ
22		4.8E-004	7.2E-003	3.2E-013	ϵ
23		4.8E-004		1.7E-004	ϵ
24			7.2E-003	1.7E-004	ϵ
25			3.0E-004	1.0E-006	ϵ
26				1.0E-006	ϵ
27			7.2E-003	1.4E-005	ϵ
28			7.2E-003	6.2E-008	ϵ
29		4.8E-004	7.2E-003	1.0E-006	ϵ
30			3.0E-004	ϵ	ϵ
31		4.8E-004	7.2E-003	1.8E-003	ϵ
32			9.2E-006	6.2E-008	ϵ
33		4.8E-004	7.2E-003	1.8E-003	2.4E-004
34		4.8E-004	9.2E-006		ϵ

Table 1.5 gives the p-values of C for the *bad* LCGs, still for $k = 2^e$. The p-values for X^2 and $-H$ are somewhat different, mainly because they are computed using the normal approximation, which is not so good with these small values of ν (the expected number of collisions is still $2^{2\nu-1}$). Here we have $\tilde{\nu} = -1$ and $\nu^* = 1$, so the strong rejections appear for a sample size n approximately 8 times smaller than for the good LCGs.

The p-values at ν^* are now ϵ instead of $-\epsilon$, which means that C (as well as X^2 and $-H$) is *much larger* than expected. For example, for $\nu = 1$, the number of collisions should be approximately Poisson with mean 2. But it

TABLE 1.6. Some p-values for X^2, for $t = 2$ and $k \approx n/8$, for good LCGs.

e	$\nu = 4$	$\nu = 5$	$\nu = 6$	$\nu = 7$	$\nu = 8$
12		–1.6E-008		ϵ	ϵ
13	–4.2E-003	–3.5E-010		ϵ	ϵ
14		–8.6E-004	$-\epsilon$		ϵ
15		–6.7E-003	$-\epsilon$		ϵ
16	–1.2E-003	–9.4E-003	–1.9E-015	$-\epsilon$	6.2E-003
17			–7.1E-015	$-\epsilon$	–3.0E-011
18			–1.4E-009	$-\epsilon$	$-\epsilon$
19		–4.9E-004	–9.0E-005	$-\epsilon$	$-\epsilon$
20			–3.0E-006	$-\epsilon$	$-\epsilon$
21			–7.8E-005	$-\epsilon$	$-\epsilon$
22		–9.9E-003		$-\epsilon$	$-\epsilon$
23				–9.9E-009	$-\epsilon$
24				–1.6E-006	$-\epsilon$
25			–5.0E-003	–1.1E-007	$-\epsilon$
26			–1.1E-003	–1.0E-005	$-\epsilon$
27					$-\epsilon$
28				–1.7E-004	–1.2E-014

is actually equal to 28 for $e = 14$, 35 for $e = 15$, 25 for $e = 16$, and so on. In other words, the points are concentrated in much fewer boxes than they should. This is due to the fact that they all lie on a small number of lines in the unit square, as indicated by the bad spectral test values.

With $k \approx \rho/8$ and $k \approx 8\rho$, the behavior is similar to that with $k \approx \rho$ and the values of ν^* are the same or slightly larger, both for the good and for the bad LCGs. The p-values are generally not as close to 0 or 1, and they get farther away from 0 or 1 when k gets farther away from ρ.

For the results that we just gave, the number of points is much less than the number of boxes. But in the classical application of the serial test [Knu81, Fis96], the number of points is usually at least 5 to 10 times the number of boxes. This is the dense case. In Tables 1.6 and 1.7, we give the p-values for X^2, for the good and bad LCGs, respectively, for $k' = n/8$. The results for $-H$ are similar. Here, X^2 follows approximately the chi-square distribution. In contrast with the previous tables, ν^* increases with e. To obtain $\pm\epsilon$, one needs a much larger n in this dense case than in the sparse case, especially when ρ is large.

To have a better view of when the $\pm\epsilon$ start to occur in this case, let us change φ. Table 1.8 gives the p-values as in Table 1.6, but with $\varphi(e) = 2^{2e/3} \approx \rho^{2/3}$. The entries marked "—" are cases where we decided not to perform the tests because n was deemed too large. Now, we have $\tilde{\nu} = 3$ and $\nu^* = 4$ for the good LCGs. The *rate* of increase of n as a function of ρ to get the small p-values is larger in the dense case than in the sparse case: $O(\rho^{2/3})$ instead of $O(\rho^{1/2})$ (empirically). For the bad LCGs, the behavior is similar and the values of $\tilde{\nu}$ and ν^* are the same.

What we have seen for $t = 2$ can also be observed in a larger number of dimensions, with slightly different values of $\tilde{\nu}$ and ν^*. Table 1.9 gives the p-values of X^2 for the good LCGs, for $k' = 2^e \approx \rho$ and $t = 8$. Here, $\nu^* = 6$, two units higher than for $t = 2$. The results for $-H$ and C are almost the same.

Table 1.10 gives the p-values of X^2 for the same parameters as in Table 1.9, but for the *overlapping* case. Again, the results for $-H$ and C are almost the same. One has the same ν^* as in the *non-overlapping* case, but the overlapping test is more economical because it requires t times less random numbers. A similar behavior can be observed for other values of t and other definitions of k'.

6.2 Serial Tests of Equidistribution for Nonlinear Generators

Table 1.12 and 1.13 show the p-values of X^2 for $t = 2$, $\varphi(e) = 2^e \approx \rho$, and $k' = 2^e$, for the inversive and cubic generators, respectively. Clearly, the sample size n required for rejection increases as $O(\rho)$ in both cases. In fact, suspect p values occur systematically only for n larger than the period length for the inversive generators, and n approximately equal to the period length for the cubic generators. Observe that when $n = \rho$, $2n$ random numbers are used in the test, so the same random numbers are used twice. But the n points are still distinct because ρ is odd. The latter holds in general provided that $\gcd(t, \rho) = 1$. The behavior is similar for $k' = 2^{e-3}$ and $k' = 2^{e+3}$, and also for C, whereas ν^* is sometimes one or two units larger for $-H$.

6.3 A Summary of the Serial Tests Results

Table 1.14 summarizes the values of $\tilde{\nu}$ and ν^* for the serial tests under different setups, for the LCGs. Table 1.15 does the same for the explicit inversive and the cubic generators. The values in parentheses lie in areas where the normal approximation is not very good. In these areas, the tests based on C (with the Poisson approximation) are more reliable. As expected, the bad LCGs fail with much smaller sample sizes than the good LCGs in dimension 2. These LCGs were chosen to be bad in two dimensions, but they turn out to be not so bad in larger dimensions (in general). In 8 dimensions, they do practically as well as the good ones in terms of $\tilde{\nu}$ and ν^*.

TABLE 1.7. Some p-values for X^2, for $t = 2$ and $k \approx n/8$, for bad LCGs.

e	$\nu = 2$	$\nu = 3$	$\nu = 4$	$\nu = 5$	$\nu = 6$
14		ϵ	ϵ	ϵ	ϵ
15		ϵ	ϵ	ϵ	ϵ
16	ϵ	4.6E-005	ϵ	ϵ	ϵ
17	8.9E-007		ϵ	ϵ	ϵ
18			ϵ	ϵ	ϵ
19		ϵ		ϵ	ϵ
20		ϵ	ϵ	ϵ	ϵ
21			ϵ		ϵ
22	6.3E-005	ϵ	ϵ	ϵ	ϵ
23				ϵ	−5.4E-005
24				ϵ	ϵ
25					ϵ
26			ϵ	ϵ	ϵ
27			8.3E-003		5.9E-004
28					
29					4.5E-003
30					
31					

TABLE 1.8. Some p-values for X^2, for $t = 2$ and $k \approx n/8$ and $\varphi(e) = 2^{2e/3}$, for good LCGs.

e	$\nu = 0$	$\nu = 1$	$\nu = 2$	$\nu = 3$	$\nu = 4$
12				−1.5E-008	
13				−1.3E-015	
14				−4.8E-011	−4.3E-007
15				−9.3E-010	$-\epsilon$
16				−2.2E-010	$-\epsilon$
17				−1.2E-009	$-\epsilon$
18				−1.4E-009	$-\epsilon$
19			−1.2E-004	−1.6E-006	$-\epsilon$
20		−3.4E-003	−4.1E-003	−2.1E-010	$-\epsilon$
21			−3.3E-003	−1.1E-006	$-\epsilon$
22				−8.9E-008	$-\epsilon$
23			−4.8E-003	−1.2E-004	$-\epsilon$
24				−2.2E-008	$-\epsilon$
25				−3.1E-009	$-\epsilon$
26			−5.5E-004	−4.0E-007	$-\epsilon$
27				−4.0E-006	—
28				−4.3E-007	—

TABLE 1.9. The p-values of X^2 for $t = 8$ and $k \approx \rho$, for good LCGs.

e	$\nu = 2$	$\nu = 3$	$\nu = 4$	$\nu = 5$	$\nu = 6$
12		–7.1E-003	–2.8E-004	ϵ	ϵ
13	5.4E-003			–1.4E-005	ϵ
14				–1.3E-007	ϵ
15		–6.3E-003	–1.9E-008	–5.2E-013	ϵ
16			–1.0E-004	–1.0E-010	ϵ
17			–8.8E-003	–3.1E-010	$-\epsilon$
18			–1.3E-003	–9.0E-012	$-\epsilon$
19		–1.2E-003		–1.2E-010	$-\epsilon$
20				–2.4E-011	$-\epsilon$
21			–9.7E-008	–4.8E-014	$-\epsilon$
22			–3.4E-003	–6.8E-012	$-\epsilon$
23			–4.9E-003	–3.1E-007	$-\epsilon$
24			–9.9E-004	–7.7E-009	$-\epsilon$
25			–1.7E-004	–4.2E-009	$-\epsilon$
26			–1.3E-004	$-\epsilon$	$-\epsilon$
27			–5.0E-003	–1.9E-010	$-\epsilon$
28			–3.1E-004	–5.6E-013	$-\epsilon$
29			–1.0E-005	$-\epsilon$	$-\epsilon$
30			–4.3E-008	–2.3E-011	$-\epsilon$
31			–6.2E-006	–1.5E-013	$-\epsilon$
32			–3.7E-007	–2.7E-015	$-\epsilon$

TABLE 1.10. The p-values of X^2 for $t = 8$ (with overlapping) and $k \approx \rho$, for good LCGs.

e	$\nu = 3$	$\nu = 4$	$\nu = 5$	$\nu = 6$
12		–2.1E-003	–6.6E-004	–8.4E-013
13				–4.8E-008
14			–1.4E-005	$-\epsilon$
15			–5.2E-008	$-\epsilon$
16			–1.2E-005	–4.3E-012
17			–3.8E-005	$-\epsilon$
18		–7.9E-004		–1.1E-011
19		–4.1E-005	–8.2E-008	$-\epsilon$
20		–2.0E-003	–1.7E-007	–1.4E-013
21		–3.6E-004	–3.8E-010	$-\epsilon$
22	–7.7E-004		–2.4E-007	$-\epsilon$
23		–1.1E-003	–4.2E-004	$-\epsilon$
24			–3.3E-010	$-\epsilon$
25		–1.6E-003	–8.4E-011	$-\epsilon$
26		–1.6E-003	–4.8E-012	$-\epsilon$
27			–7.0E-004	$-\epsilon$
28		–1.5E-003	–3.8E-009	$-\epsilon$
29	–1.9E-003	–1.7E-009	–1.2E-013	$-\epsilon$
30		–5.5E-006	–6.2E-011	$-\epsilon$
31		–1.3E-005	–1.4E-014	$-\epsilon$
32		–9.4E-006	–1.8E-009	$-\epsilon$

TABLE 1.11. The p-values of X^2 for $t = 8$ (with overlapping) and $k \approx \rho$, for bad LCGs.

e	$\nu = 3$	$\nu = 4$	$\nu = 5$	$\nu = 6$	$\nu = 7$
12				2.9E-003	ϵ
13		–1.5E-003	–1.1E-005	–3.2E-013	ϵ
14					8.0E-003
15	7.0E-003		1.1E-003	ϵ	ϵ
16					
17		–9.1E-003	–4.6E-013	$-\epsilon$	$-\epsilon$
18					–3.7E-003
19			–1.9E-004	–2.9E-014	$-\epsilon$
20				–1.6E-003	–7.8E-004
21					
22		–7.9E-003	–8.7E-010	$-\epsilon$	$-\epsilon$
23				–3.6E-003	–3.2E-008
24				–1.1E-003	$-\epsilon$
25		–2.6E-004	–4.2E-008	$-\epsilon$	$-\epsilon$
26		–1.4E-003	–3.1E-005	$-\epsilon$	$-\epsilon$
27			–4.9E-004	–1.6E-007	$-\epsilon$
28			1.7E-005	ϵ	ϵ
29				–8.4E-003	–6.9E-008
30			–4.4E-004	–9.3E-010	$-\epsilon$
31		–3.2E-003	–6.6E-005	–1.0E-013	—
32			–2.0E-006	$-\epsilon$	—

TABLE 1.12. The p-values of X^2 for $t = 2$ and $k \approx \rho$, for inversive generators.

e	$\nu = -1$	$\nu = 0$	$\nu = 1$	$\nu = 2$
10			ϵ	ϵ
11			ϵ	ϵ
12			ϵ	ϵ
13	–3.9E-003		ϵ	ϵ
14			ϵ	ϵ
15			ϵ	ϵ
16	–7.1E-003	–2.4E-003	ϵ	ϵ
17		–3.0E-004	ϵ	ϵ
18			ϵ	ϵ
19			ϵ	ϵ
20			ϵ	ϵ
21		1.2E-003	ϵ	—
22			—	—

TABLE 1.13. The p-values of X^2 for $t = 2$ and $k \approx \rho$, for cubic generators.

e	$\nu = -1$	$\nu = 0$	$\nu = 1$
14		ϵ	ϵ
15		ϵ	ϵ
16		ϵ	ϵ
17		ϵ	ϵ
18		ϵ	ϵ
19		ϵ	ϵ
20		ϵ	ϵ
21		ϵ	ϵ
22		ϵ	—

TABLE 1.14. The good and bad LCGs in the serial tests, with $\varphi(e) = 2^{e/2}$.

Overlap	t	k'	Test	Good LCGs		Bad LCGs	
				$\tilde{\nu}$	ν^*	$\tilde{\nu}$	ν^*
No	2	2^e	X^2	(2)	4	(−1)	(0)
"	2	2^e	$-H$	(2)	4	(−1)	(0)
"	2	2^e	C	2	4	−1	1
"	2	2^{e+3}	X^2	(4)	5	(0)	(0)
"	2	2^{e+3}	$-H$	(4)	5	(0)	(0)
"	2	2^{e+3}	C	4	5	0	2
"	2	2^{e-3}	X^2	4	6	(−1)	(0)
"	2	2^{e-3}	$-H$	4	6	(−1)	(0)
"	2	2^{e-3}	C	4	6	−1	1
"	2	2^{e-5}	X^2	5	7	(−1)	(0)
"	2	2^{e-5}	$-H$	5	7	(−1)	(0)
"	2	2^{e-5}	C	5	7	−1	1
"	8	2^e	X^2	4	6	5	6
"	8	2^e	$-H$	4	6	5	6
"	8	2^e	C	4	6	5	6
"	8	2^{e+3}	X^2	4	6	5	6
"	8	2^{e+3}	$-H$	4	6	5	6
"	8	2^{e+3}	C	4	6	5	6
"	8	2^{e-3}	X^2	5	6	5	7
"	8	2^{e-3}	$-H$	5	7	5	7
"	8	2^{e-3}	C	5	6	5	7
Yes	8	2^e	X^2	4	6	5	7
"	8	2^e	C	4	6	5	7
"	8	2^{e+3}	X^2	5	6	5	7
"	8	2^{e+3}	C	5	6	5	6
"	8	2^{e-3}	X^2	5	7	6	8

TABLE 1.15. The inversive and cubic generators in the serial tests, with $\varphi(e) = 2^e$.

Overlap	t	k'	Test	Inversive $\tilde{\nu}$	Inversive ν^*	Cubic $\tilde{\nu}$	Cubic ν^*
No	2	2^e	X^2	1	1	0	0
"	2	2^e	$-H$	2	3	0	0
"	2	2^e	C	1	1	0	0
"	2	2^{e+3}	X^2	1	1	0	0
"	2	2^{e+3}	$-H$	1	1	0	0
"	2	2^{e+3}	C	1	1	0	0
"	2	2^{e-3}	X^2	1	1	0	0
"	2	2^{e-3}	$-H$	1	1	1	1
"	8	2^e	X^2	1	1	0	0
"	8	2^e	$-H$	2	3	0	0
"	8	2^e	C	1	1	0	0
"	8	2^{e+3}	X^2	0	1	0	0
"	8	2^{e+3}	$-H$	1	1	0	0
"	8	2^{e+3}	C	1	1	0	0
"	8	2^{e-3}	X^2	1	1	0	0
"	8	2^{e-3}	$-H$	1	1	1	1
Yes	8	2^e	X^2	1	1	0	0
"	8	2^{e+3}	X^2	1	1	0	0
"	8	2^{e-3}	X^2	1	1	0	0

TABLE 1.16. NP test with $t = 2$ and $N = 1$ for good LCGs.

e	$\nu = 0$	$\nu = 1$	$\nu = 2$	$\nu = 3$
14			5.5E-007	ϵ
15		4.5E-003	3.9E-010	ϵ
16		7.6E-003	3.3E-009	ϵ
17			4.7E-008	ϵ
18			7.5E-006	ϵ
19		2.9E-003	7.1E-011	ϵ
20		1.8E-003	1.0E-011	ϵ
21			1.9E-008	ϵ
22		9.5E-003	8.1E-009	ϵ
23		1.7E-003	9.2E-012	ϵ
24		7.0E-004	2.3E-013	ϵ
25			1.4E-008	ϵ
26		2.6E-003	4.6E-011	ϵ
27		8.8E-003	5.9E-009	ϵ
28			1.0E-007	ϵ
29		6.6E-004	1.9E-013	ϵ
30			1.8E-007	ϵ
31		1.2E-003	2.1E-012	ϵ
32			6.5E-008	ϵ
33			2.0E-006	ϵ
34		3.0E-003	7.6E-011	ϵ
35		8.1E-003	4.4E-009	ϵ
36		7.2E-003	2.8E-009	ϵ

TABLE 1.17. The 32-NP test with $t = 2$ and $N = 1$ for good LCGs.

e	$\nu = -2$	$\nu = -1$	$\nu = 0$	$\nu = 1$
22	—	—	—	ϵ
24	—	—	ϵ	ϵ
26	—	4.3E–09	ϵ	ϵ
28		2.9E–04	ϵ	ϵ
30	5.9E–05	ϵ	ϵ	ϵ
32	3.4E–04	ϵ	ϵ	ϵ
34	1.6E–03	3.1E–03	ϵ	ϵ
36			ϵ	ϵ
38		6.1E–04	ϵ	ϵ
40		8.8E–12	ϵ	ϵ

TABLE 1.18. The NP test with $t = 2$ and $N = 32$ for good LCGs.

e	$\nu = -2$	$\nu = -1$	$\nu = 0$	$\nu = 1$
14			3.5E–07	ϵ
16			4.4E–09	ϵ
18			2.3E–04	ϵ
20			4.4E–13	ϵ
22			6.1E–10	ϵ
24			ϵ	ϵ
26			9.9E–12	ϵ
28			4.0E–07	ϵ
30			1.9E–06	ϵ
32			5.2E–07	ϵ
34		2.4E–03	2.7E–11	ϵ
36			1.9E–09	ϵ
38			1.3E–11	ϵ
40		3.2E–03	3.0E–13	ϵ

TABLE 1.19. The NP-S test with $t = 2$ and $N = 32$ for good LCGs.

e	$\nu = -4$	$\nu = -3$	$\nu = -2$	$\nu = -1$
14	—	—	ϵ	ϵ
16	—	2.8E–03	ϵ	ϵ
18	7.5E–03	5.4E–12	ϵ	ϵ
20		2.9E–03	ϵ	ϵ
22		1.1E–04	ϵ	ϵ
24		1.8E–15	ϵ	ϵ
26		ϵ	ϵ	ϵ
28			ϵ	ϵ
30		1.5E–04	ϵ	ϵ
32	2.1E–05	5.5E–04	ϵ	ϵ
34		6.2E–09	ϵ	ϵ
36		1.6E–14	ϵ	ϵ
38		8.7E–04	ϵ	ϵ
40	9.5E–03		ϵ	ϵ

TABLE 1.20. NP test with $t = 2$ and $N = 1$ for bad LCGs.

e	$\nu = 4$	$\nu = 5$	$\nu = 6$	$\nu = 7$	$\nu = 8$
14			$-\epsilon$	$-\epsilon$	$-\epsilon$
15		1.9E-003	1.4E-011	$-\epsilon$	$-\epsilon$
16		1.2E-004	ϵ	$-\epsilon$	$-\epsilon$
17		3.6E-003	1.6E-010	ϵ	$-\epsilon$
18		5.1E-003	6.7E-010	ϵ	$-\epsilon$
19		2.6E-003	4.8E-011	ϵ	ϵ
20		3.4E-003	1.3E-010	ϵ	ϵ
21		6.3E-003	1.6E-009	ϵ	ϵ
22		2.7E-003	5.5E-011	ϵ	ϵ
23		2.8E-003	6.5E-011	ϵ	ϵ
24		4.8E-003	5.5E-010	ϵ	ϵ
25		8.9E-003	6.4E-009	ϵ	ϵ
26		2.8E-003	6.1E-011	ϵ	ϵ

6.4 Close-Pairs Tests for LCGs

We applied the close-pairs tests to the good and bad LCGs, with $\varphi(e) = 2^{e/2}$. We now give the results for $t = 2$ and $p = \infty$. Tables 1.16 to 1.19 give the p-values for the NP test with $N = 1$, the 32-NP test with $N = 1$, the NP test with $N = 32$, and the NP-S test with $N = 32$, for the good LCGs. Tables 1.20 gives the results of the NP test with $N = 1$, for the bad LCGs. The entries marked "—" are those for which the test was not run because n was deemed either too small for the Poisson process approximation to be good, or too large. A summary of the other results is given in Table 1.25.

For the NP tests with $t = 2$ and $N = 1$, the good LCGs fail much *sooner* than the bad ones. Why? The LCGs normally fail the NP test because the distance between the two nearest points is *too large*. The reason is that this distance has a lower bound determined by the lattice structure of the point set Ψ_2. And this lower bound turns out to be *smaller* when all the points lie on a set of lines which are *farther* apart, because the successive points on a given line are then closer to each other. The so-called bad LCGs then perform better with respect to the NP test with $N = 1$ since their lower bound is smaller.

For the M-NP tests with $M > 1$ or for the NP-S or NP-PR tests with $N > 1$, the good and bad LCGs fail at about the same sample sizes. For the bad LCGs, the minimal distance tends to be smaller, but a clustering between the values of the few shortest distances occurs anyhow, and this is detected by the AD statistic just as well as for the good LCGs. For the NP test with $N = 32$ (without the spacings or power ratio transformations), the bad LCGs fail significantly sooner than the good ones. The effect of the transformations does not show up in the value of ν^* for the bad LCGs, but the transformations are nevertheless effective (although less spectacularly than for the good LCGs) and this shows up when one looks more closely

TABLE 1.21. The 32-NP test with $t = 8$ and $N = 1$ for good LCGs.

e	$\nu = -2$	$\nu = -1$	$\nu = 0$	$\nu = 1$
24	—	—	ϵ	ϵ
25	—	—	ϵ	ϵ
26	—	5.5E-014	ϵ	ϵ
27	—	5.4E-003	ϵ	ϵ
28		ϵ	ϵ	ϵ
29		ϵ	ϵ	ϵ
30		ϵ	ϵ	ϵ
31	5.3E-003	5.5E-013	ϵ	ϵ
32		7.9E-003	ϵ	ϵ
33	1.8E-006	1.3E-012	ϵ	ϵ
34		2.7E-007	ϵ	ϵ
35		ϵ	ϵ	ϵ
36		6.1E-011	ϵ	ϵ

at the values of the AD statistics.

We also tried close-pairs tests in more than 2 dimensions. As an illustration, Tables 1.21 and 1.22 give the results for the good and bad LCGs, for the 32-NP test in 8 dimensions. These two classes of generators perform about the same with respect to this test. In fact, in 8 dimensions, the bad LCGs are not so bad as in 2 dimensions (they were actually constructed to be bad in 2 dimensions) and this shows up in the results. This remains true in general for the other close-pairs tests and for other values of $t > 2$.

6.5 Close-Pairs Tests for Nonlinear Generators

The behavior of the inversive and cubic generators with the close-pairs tests is similar to the behavior of these generators with the serial tests. The sample size n required for rejection increases as $O(\rho)$, not as $O(\sqrt{\rho})$. As illustrations, Table 1.23 and Table 1.24 give the p-values for the inversive generators, for $t = 2$, $\varphi(e) = 2^e \approx \rho$, for the NP and 32-NP tests, respectively. In both cases, the test is single-level ($N = 1$). The 32-NP test starts detecting something when the sample size n reaches approximately $1/64$ or $1/32$ of the period length, whereas the NP test needs n approximately equal to the period length. The behavior of the compound cubic generators is similar. The results are summarized in the next subsection.

6.6 A Summary of the Close-Pairs Tests Results

Table 1.25 summarizes the values of $\tilde{\nu}$ and ν^* for the close-pairs tests in two dimensions.

In two dimensions, for $N = 1$, the NP test suggested by [Rip87] needs $n \approx 8\sqrt{\rho}$ for the good LCGs and $n \approx 128\sqrt{\rho}$ for the bad LCGs, whereas

TABLE 1.22. The 32-NP test with $t = 8$ and $N = 1$ for bad LCGs.

e	$\nu = -2$	$\nu = -1$	$\nu = 0$	$\nu = 1$
24	—	—	ϵ	ϵ
25	—	—	ϵ	ϵ
26	—	ϵ	ϵ	ϵ
27	—	5.9E-005	ϵ	ϵ
28		3.7E-005	ϵ	ϵ
29	1.1E-005	2.4E-008	ϵ	ϵ
30		ϵ	ϵ	ϵ
31		5.3E-006	ϵ	ϵ
32		6.6E-011	ϵ	ϵ
33		4.9E-012	ϵ	ϵ
34		ϵ	ϵ	ϵ
35	1.2E-005	ϵ	ϵ	ϵ
36		1.4E-011	ϵ	ϵ

TABLE 1.23. The p-values for the NP test with $t = 2$, for inversive generators.

e	$\nu = -3$	$\nu = -2$	$\nu = -1$	$\nu = 0$	$\nu = 1$
10				$-\epsilon$	$-\epsilon$
11				$-\epsilon$	$-\epsilon$
12				$-\epsilon$	$-\epsilon$
13				$-\epsilon$	$-\epsilon$
14				$-\epsilon$	$-\epsilon$
15				$-\epsilon$	$-\epsilon$
16				$-\epsilon$	$-\epsilon$
17				$-\epsilon$	$-\epsilon$
18				$-\epsilon$	$-\epsilon$
19				$-\epsilon$	$-\epsilon$
20				$-\epsilon$	$-\epsilon$
21				$-\epsilon$	$-\epsilon$
22				$-\epsilon$	—

TABLE 1.24. The p-values for the 32-NP test with $t = 2$, for inversive generators.

e	$\nu = -8$	$\nu = -7$	$\nu = -6$	$\nu = -5$	$\nu = -4$
17	—	—	—	ϵ	1.5E-009
18	—	—		5.7E-007	7.7E-015
19	—	2.8E-004	1.3E-004	ϵ	ϵ
20			1.7E-007	1.0E-009	8.2E-012
21					ϵ
22		4.8E-005		2.1E-007	ϵ
23			7.5E-006	2.1E-014	ϵ
24				2.5E-013	ϵ
25		1.4E-003		ϵ	ϵ
26			7.8E-006	1.4E-013	ϵ
27				7.9E-008	—

TABLE 1.25. Values of $\tilde{\nu}$ and ν^* for the close-pairs tests for $p = \infty$ and $t = 2$.

Generators	Test	N	$\tilde{\nu}$	ν^*	$\nu^* + \log_2 N$
Good LCGs	NP	1	2	3	3
”	32-NP	1	−1	0	0
”	NP	16	1	2	6
”	NP-S	16	−2	1	5
”	NP-PR	16	−2	−1	3
”	NP	32	0	1	6
”	NP-S	32	−3	−2	3
”	NP-PR	32	−3	−2	3
”	16-NP	16	−2	−1	3
”	32-NP	32	−2	−1	4
Bad LCGs	NP	1	6	7	7
”	32-NP	1	−1	0	0
”	NP	32	−3	−2	3
”	NP-S	32	−4	−2	3
”	NP-PR	32	−4	−2	3
Inversive	NP	1	0	0	0
”	32-NP	1	−5	−4	−4
”	NP-S	32	−8	−7	−2
”	NP-PR	32	−8	−7	−2
Cubic	NP	1	0	0	0
”	32-NP	1	−5	−3	−3
”	NP-S	32	−7	−6	−1
”	NP-PR	32	−7	−6	−1

TABLE 1.26. Values of $\tilde{\nu}$ and ν^* for some close-pairs tests for $p = \infty$ and $t = 8$.

Generators	Test	N	$\tilde{\nu}$	ν^*	$\nu^* + \log_2 N$
Good LCGs	NP	1	1	2	2
"	32-NP	1	−1	0	0
Bad LCGs	NP	1	2	3	3
"	32-NP	1	−1	0	0
Inversive	NP	1	0	0	0
"	32-NP	1	−1	0	0
Cubic	NP	1	0	0	0
"	32-NP	1	0	0	0

the 32-NP test needs $n \approx \sqrt{\rho}$ in both cases. With $N = 32$, the NP-S and NP-PR tests need a sample size approximately 8 times smaller than the NP test for the good LCGs. The spacings and power ratio transformations really help. The inversive and cubic generators start failing at sample sizes equal to some value around $\rho/32$ or $\rho/16$ in 2 dimensions.

In 8 dimensions, the good and bad LCGs perform almost equally well. For the inversive and cubic generators, the numbers in the table do not tell the whole story, because there is a lot of variation between the individual generators (i.e., the different values of e) in the results of the 32-NP test. The table gives $\nu^* = 0$ but some p-values around 10^{-15} or less have been observed at $\nu = -3$ or -2, for some values of e.

7 How do Real-Life Generators Fare in These Tests?

The results of the systematic experiments of the previous sections can be used to predict the sample size at which a given generator of the type considered will start to fail a given test. For example, any full-period LCG with modulus 2^{32} or less should fail unequivocally a 32-NP test in two dimensions with sample size $n = 2^{16}$ or more. And any LCG with period length 2^{48} should fail the 32-NP test with $n = 2^{24}$. These expectations are confirmed in [LCS97], where some widely used LCGs have been submitted to several close-pairs tests. Similar experiments with the serial tests are reported in [LSW98, LW97]. We can conclude from this that if an LCG is to be implemented in a general-purpose statistical or simulation package, its period length should then be *at least* larger than the square of the maximal number of output values that is expected to be used in any simulation or statistical experiment.

For an equivalent period length, explicit inversive and cubic generators fare better in the tests than the LCGs, but the inversive ones are much slower, whereas the cubic ones are not yet well-understood from the theoretical viewpoint and few parameters are available for full-period instances. In our opinion, the importance of these generators lies more in the fact that there are no lattice structures than in the irregularity of the full-period point sets with respect to discrepancy.

Other generators described in [L'E96a, L'E97a, MN98] have so far done well in all the statistical tests that have been applied to them. These generators are also reasonably fast and can be recommended for general use.

8 Acknowledgments

This work has been supported by NSERC-Canada grants # OGP0110050 and # SMF0169893, and FCAR-Québec grant # 93-ER-1654 to the first author and by FWF-Austria grants P9285, P11143, and P12654 to the second author. This paper was written while the first author was visiting the pLab group headed by the second author at Salzburg University. Richard Simard helped in the implementations of the tests.

9 References

[AG85] L. Afflerbach and H. Grothe. Calculation of Minkowski-reduced lattice bases. *Computing*, 35:269–276, 1985.

[AGG89] R. Arratia, L. Goldstein, and L. Gordon. Two moments suffice for Poisson approximation: The Chen-Stein method. *The Annals of Probability*, 17:9–25, 1989.

[Alt88] N. S. Altman. Bit-wise behavior of random number generators. *SIAM Journal on Scientific and Statistical Computing*, 9(5):941–949, 1988.

[CLT93] R. Couture, P. L'Ecuyer, and S. Tezuka. On the distribution of k-dimensional vectors for simple and combined Tausworthe sequences. *Mathematics of Computation*, 60(202):749–761, S11–S16, 1993.

[Com91] A. Compagner. The hierarchy of correlations in random binary sequences. *Journal of Statistical Physics*, 63:883–896, 1991.

[Com95] A. Compagner. Operational conditions for random number generation. *Physical Review E*, 52(5-B):5634–5645, 1995.

[CS88] J. H. Conway and N. J. A. Sloane. *Sphere Packings, Lattices and Groups*. Grundlehren der Mathematischen Wissenschaften 290. Springer-Verlag, New York, 1988.

[DR81] E. J. Dudewicz and T. G. Ralley. *The Handbook of Random Number Generation and Testing with TESTRAND Computer Code.* American Sciences Press, Columbus, Ohio, 1981.

[Dur73] J. Durbin. *Distribution Theory for Tests Based on the Sample Distribution Function.* SIAM CBMS-NSF Regional Conference Series in Applied Mathematics. SIAM, Philadelphia, 1973.

[DMST95] E. J. Dudewicz, E. C. van der Meulen, M. G. SriRam, and N. K. W. Teoh. Entropy-based random number evaluation. *American Journal of Mathematical and Management Sciences*, 15:115–153, 1995.

[EH97] J. Eichenauer-Herrmann and E. Herrmann. Compound cubic congruential pseudorandom numbers. *Computing*, 59:85–90, 1997.

[EHW97] J. Eichenauer-Herrmann, E. Herrmann, and S. Wegenkittl. A survey of quadratic and inversive congruential pseudorandom numbers. In P. Hellekalek, G. Larcher, H. Niederreiter, and P. Zinterhof, editors, *Monte Carlo and Quasi-Monte Carlo Methods in Scientific Computing*, volume 127 of *Lecture Notes in Statistics*, pages 66–97, New York, 1997. Springer-Verlag.

[Eic92] J. Eichenauer-Herrmann. Inversive congruential pseudorandom numbers: A tutorial. *International Statistical Reviews*, 60:167–176, 1992.

[Fis90] G. S. Fishman. Multiplicative congruential random number generators with modulus 2^β: An exhaustive analysis for $\beta = 32$ and a partial analysis for $\beta = 48$. *Mathematics of Computation*, 54(189):331–344, Jan 1990.

[Fis96] G. S. Fishman. *Monte Carlo: Concepts, Algorithms, and Applications.* Springer Series in Operations Research. Springer-Verlag, New York, 1996.

[FM86] G. S. Fishman and L. S. Moore III. An exhaustive analysis of multiplicative congruential random number generators with modulus $2^{31} - 1$. *SIAM Journal on Scientific and Statistical Computing*, 7(1):24–45, 1986.

[Goo53] I. J. Good. The serial test for sampling numbers and other tests for randomness. *Proceedings of the Cambridge Philos. Society*, 49:276–284, 1953.

[Hel95] P. Hellekalek. Inversive pseudorandom number generators: Concepts, results, and links. In C. Alexopoulos, K. Kang, W. R.

Lilegdon, and D. Goldsman, editors, *Proceedings of the 1995 Winter Simulation Conference*, pages 255–262. IEEE Press, 1995.

[Hel97] P. Hellekalek. On correlation analysis of pseudorandom numbers. In H. Niederreiter, P. Hellekalek, G. Larcher, and P. Zinterhof, editors, *Monte Carlo and Quasi-Monte Carlo Methods 1996*, volume 127 of *Lecture Notes in Statistics*, pages 251–265. Springer-Verlag, New York, 1997.

[HL97] P. Hellekalek and H. Leeb. Dyadic diaphony. *Acta Arithmetica*, 80:187–196, 1997.

[HN98] P. Hellekalek and H. Niederreiter. The weighted spectral test: Diaphony. *ACM Transactions on Modeling and Computer Simulation*, 8:43–60, 1998.

[Hol72] L. Holst. Asymptotic normality and efficiency for certain goodness-of-fit tests. *Biometrika*, 59(1):137–145, 1972.

[JHK96] F. James, J. Hoogland, and R. Kleiss. Multidimensional sampling for simulation and integration: Measures, discrepancies, and quasi-random numbers. Submitted to Computer Physics Communications, 1996.

[Knu81] D. E. Knuth. *The Art of Computer Programming, Volume 2: Seminumerical Algorithms.* Addison-Wesley, Reading, Mass., second edition, 1981.

[LBC93] P. L'Ecuyer, F. Blouin, and R. Couture. A search for good multiple recursive random number generators. *ACM Transactions on Modeling and Computer Simulation*, 3(2):87–98, 1993.

[LCC96] P. L'Ecuyer, A. Compagner, and J.-F. Cordeau. Entropy tests for random number generators. Submitted, 1996.

[LCS97] P. L'Ecuyer, J.-F. Cordeau, and R. Simard. Close-point spatial tests and their application to random number generators. Submitted, 1997.

[L'E92] P. L'Ecuyer. Testing random number generators. In *Proceedings of the 1992 Winter Simulation Conference*, pages 305–313. IEEE Press, Dec 1992.

[L'E94] P. L'Ecuyer. Uniform random number generation. *Annals of Operations Research*, 53:77–120, 1994.

[L'E96a] P. L'Ecuyer. Combined multiple recursive random number generators. *Operations Research*, 44(5):816–822, 1996.

[L'E96b] P. L'Ecuyer. Maximally equidistributed combined Tausworthe generators. *Mathematics of Computation*, 65(213):203–213, 1996.

[L'E97a] P. L'Ecuyer. Good parameters and implementations for combined multiple recursive random number generators. Manuscript, 1997.

[L'E97b] P. L'Ecuyer. Random number generators and empirical tests. In P. Hellekalek, G. Larcher, H. Niederreiter, and P. Zinterhof, editors, *Monte Carlo and Quasi-Monte Carlo Methods in Scientific Computing*, volume 127 of *Lecture Notes in Statistics*, pages 124–138. Springer-Verlag, New York, 1997.

[L'E97c] P. L'Ecuyer. Tests based on sum-functions of spacings for uniform random numbers. *Journal of Statistical Computation and Simulation*, 59:251–269, 1997.

[L'E98a] P. L'Ecuyer. Random number generation. In Jerry Banks, editor, *Handbook on Simulation*. Wiley, 1998. To appear. Also GERAD technical report number G-96-38.

[L'E98b] P. L'Ecuyer. A table of linear congruential generators of different sizes and good lattice structure. *Mathematics of Computation*, 1998. To appear.

[L'E98c] P. L'Ecuyer. Tables of maximally equidistributed combined LFSR generators. *Mathematics of Computation*, 1998. To appear.

[L'E00] P. L'Ecuyer. TestU01: Un logiciel pour appliquer des tests statistiques à des générateurs de valeurs aléatoires. In preparation, Circa 2000.

[Lee97] H. Leeb. *Stochastic Properties of Diaphony*. PhD thesis, Department of Mathematics, University of Salzburg, Austria, November 1997.

[LH98] H. Leeb and P. Hellekalek. Strong and weak laws for the spectral test and related quantities. In preparation, 1998.

[LSW98] P. L'Ecuyer, R. Simard, and S. Wegenkittl. Sparse serial tests of randomness. In preparation, 1998.

[LW97] H. Leeb and S. Wegenkittl. Inversive and linear congruential pseudorandom number generators in empirical tests. *ACM Transactions on Modeling and Computer Simulation*, 7(2):272–286, 1997.

[Mar85] G. Marsaglia. A current view of random number generators. In *Computer Science and Statistics, Sixteenth Symposium on the Interface*, pages 3–10, North-Holland, Amsterdam, 1985. Elsevier Science Publishers.

[Mar96] G. Marsaglia. DIEHARD: a battery of tests of randomness. See `http://stat.fsu.edu/~geo/diehard.html`, 1996.

[MK94] M. Matsumoto and Y. Kurita. Twisted GFSR generators II. *ACM Transactions on Modeling and Computer Simulation*, 4(3):254–266, 1994.

[MK96] M. Matsumoto and Y. Kurita. Strong deviations from randomness in m-sequences based on trinomials. *ACM Transactions on Modeling and Computer Simulation*, 6(2):99–106, 1996.

[MN98] M. Matsumoto and T. Nishimura. Mersenne twister: A 623-dimensionally equidistributed uniform pseudorandom number generator. *ACM Transactions on Modeling and Computer Simulation*, 8(1):3–30, 1998.

[Mor75] C. Morris. Central limit theorems for multinomial sums. *The Annals of Statistics*, 3:165–188, 1975.

[MZ93] G. Marsaglia and A. Zaman. Monkey tests for random number generators. *Computers Math. Applic.*, 26(9):1–10, 1993.

[Nie92] H. Niederreiter. *Random Number Generation and Quasi-Monte Carlo Methods*, volume 63 of *SIAM CBMS-NSF Regional Conference Series in Applied Mathematics*. SIAM, Philadelphia, 1992.

[PK97] S. Pincus and R. E. Kalman. Not all (possibly) “Random” sequences are created equal. *Proceedings of the National Academy of Sciences of the USA*, 94:3513–3518, 1997.

[RC88] T. R. C. Read and N. A. C. Cressie. *Goodness-of-Fit Statistics for Discrete Multivariate Data*. Springer Series in Statistics. Springer-Verlag, New York, 1988.

[Rip77] B. D. Ripley. Modelling spatial patterns. *Journal of the Royal Statistical Society, Series B*, 39:172–212, 1977.

[Rip87] B. D. Ripley. *Stochastic Simulation*. Wiley, New York, 1987.

[RS78] B. D. Ripley and B. W. Silverman. Quick tests for spatial interaction. *Biometrika*, 65(3):641–642, 1978.

[SB78] B. Silverman and T. Brown. Short distances, flat triangles and Poisson limits. *Journal of Applied Probability*, 15:815–825, 1978.

[SHB97] L. N. Shchur, J. R. Heringa, and H. W. J. Blöte. Simulation of a directed random-walk model: The effect of pseudo-random number correlations. *Physica A*, 241:579–592, 1997.

[Ste86a] M. S. Stephens. Tests based on EDF statistics. In R. B. D'Agostino and M. S. Stephens, editors, *Goodness-of-Fit Techniques*. Marcel Dekker, New York and Basel, 1986.

[Ste86b] M. S. Stephens. Tests for the uniform distribution. In R. B. D'Agostino and M. S. Stephens, editors, *Goodness-of-Fit Techniques*, pages 331–366. Marcel Dekker, New York and Basel, 1986.

[Str94] O. Strauch. L^2 discrepancy. *Math. Slovaca*, 44(5):601–632, 1994.

[Tak94] K. Takashima. Sojourn time test for maximum-length linear recurring sequences with characteristic primitive trinomials. *J. Japanese Soc. Comp. Satist.*, 7:77–87, 1994.

[Tau65] R. C. Tausworthe. Random numbers generated by linear recurrence modulo two. *Mathematics of Computation*, 19:201–209, 1965.

[Tez95] S. Tezuka. *Uniform Random Numbers: Theory and Practice.* Kluwer Academic Publishers, Norwell, Mass., 1995.

[TL91] S. Tezuka and P. L'Ecuyer. Efficient and portable combined Tausworthe random number generators. *ACM Transactions on Modeling and Computer Simulation*, 1(2):99–112, 1991.

[VANK94] I. Vattulainen, T. Ala-Nissila, and K. Kankaala. Physical tests for random numbers in simulations. *Physical Review Letters*, 73(19):2513–2516, 11 1994.

[VANK95] I. Vattulainen, T. Ala-Nissila, and K. Kankaala. Physical models as tests of randomness. *Physical Review E*, 52(3):3205–3213, 1995.

[WC93] D. Wang and A. Compagner. On the use of reducible polynomials as random number generators. *Mathematics of Computation*, 60:363–374, 1993.

[Weg95] S. Wegenkittl. Empirical testing of pseudorandom number generators. Master's thesis, University of Salzburg, 1995.

Peter Hellekalek
Institute of Mathematics, University of Salzburg
Hellbrunnerstrasse 34, A-5020 Salzburg, Austria
e-mail: peter.hellekalek@sbg.ac.at
web: http://random.mat.sbg.ac.at/team/peter.html

Pierre L'Ecuyer
Institute of Mathematics, University of Salzburg and
Département d'informatique et de recherche opérationnelle
Université de Montréal
C.P. 6128, Succ. Centre-Ville, Montréal, H3C 3J7, Canada
web: http://www.iro.umontreal.ca/~lecuyer